Optimal Power Allocation for a Renewable Energy Source

Abhinav Sinha and Prasanna Chaporkar
Electrical Engineering Department, Indian Institute of Technology, Bombay, India.
{abhinavsinha,chaporkar}@ee.iitb.ac.in

Abstract—**Battery powered transmitters face energy constraint, replenishing their energy by a renewable energy source (like solar or wind power) can lead to longer lifetime. We consider here the problem of finding the optimal power allocation under random channel conditions for a wireless transmitter, such that rate of information transfer is maximized. Here a rechargeable battery, which is periodically charged by renewable source, is used to power the transmitter. All of above is formulated as a Markov Decision Process. Structural properties like the monotonicity of the optimal value and policy derived in this paper will be of vital importance in understanding the kind of algorithms and approximations needed in real-life scenarios. The effect of curse of dimensionality which is prevalent in Dynamic programming problems can thus be reduced. We show our results under the most general of assumptions.**

Index Terms—**Optimal reward function, Monotone optimal policy, Concavity, Stochastic domination.**

I. INTRODUCTION

AS we move towards hand-held devices that use wireless transmitters, there is an exceeding need to prolong the lifetime of their batteries without having to manually recharge them on a regular basis. One natural solution to such a problem is to utilize the environment, i.e., have a renewable energy source recharge the battery periodically. This will enable the system to be self-sustaining. List of renewable energy sources include solar power, wind energy, geothermal energy and ocean energy (tidal and wave). Our objective here is to maximize the throughput of a wireless transmitter enabled with renewable energy source. (A lot of work in this regard has also been done to optimize the performance of the battery powered sensor (see Chang [1], Hou [2]) and also in field of energy-harvesting (see Yasser [3]). A recent paper has experimentally shown it possible to power a remote sensor via magnetic resonance without being in contact with the sensor, see Kurs [4])

The renewable sources of energy are better modelled as random sources due to the lack of control that we have over the source (for example in wind energy, speed of the

winds is not in our control). Thus the key challenges we face are on account of having randomness in recharge energy from the renewable source and randomness in channel state. Also since we have a battery, the maximum energy that can be stored at any point of time is limited. This is quite different in contrast to having a constraint only in terms of average power used. There could be a case for not operating at energy levels close to maximum lest added energy could go to waste. Whereas randomness in channel state could see the optimal policy conserving energy while waiting for a better channel to come. We hope to answer for such trade-off in this paper.

We model the problem of maximizing throughput of renewable energy empowered wireless transmitter as an infinite horizon discounted reward Markov Decision Process (MDP). We will use the reward function ($J^\star$), which represents the overall throughput, to compare policies. Optimal policies for us would mean deciding on what power to allocate for every possible value of battery state and channel state (defined together as *states*) so as to obtain maximum overall reward ($J^\star$) for every state. Generally MDP or dynamic programming solutions follow the "Curse of dimensionality", because the state space tends to be exponential in one or more system parameter. That is the case in our problem as well. Higher complexity solutions are not preferred as it would become a nightmare to implement it. In such a case, having some kind of structure on the solution will have big advantage implementation-wise, not to mention having more analytic tractability of the problem. Our contribution here is to prove the non-decreasing nature of the optimal policy w.r.t states. Our proofs rely only on standard results and techniques used in MDP's. Monotonicity in optimal policy is also important as it tells us about how the structure of the system is impervious to various situations like having different probability distributions on channel state and recharge energy. Once we have proven non-decreasing optimal policy, the search space automatically reduces. Moreover on the basis of this we can also try to get the threshold behaviour (approximately if suitable) which will give us

chance to make the implementation in real-time.

As far as structural properties go, monotonicity for the optimal policy is one of the most basic results. Hence there has been a plethora of work on the matter. One of the earliest method to prove monotonicity was provided by Serfozo [5]. In his book [6], Martin Puterman has provided sufficient conditions for the same as well, here however we approach the problem in a different manner (we show results based on properties of $J^\star$ rather than the Transition Probability Matrix). There has also been a lot of work on optimal policy for rechargeable sensors but with different considerations, in [7] we can find a policy which not only takes into account the rate of information transfer but also actual throughput for the queued data. Similarly, in [8], the authors have dealt with the finite horizon equivalent and have given an on-line policy which can guarantee fraction of the optimal throughput.

After defining the problem we set up the equations for finding the solution in section II. In section III we begin by proving results about monotonicity (non-decreasing) and concavity of $J^\star$ and then move on to our main result where we prove that the Optimal Power Allocation function is non-decreasing. Once we have our main structural result, we talk of possible generalizations from this framework. In section IV we present simulation results for verification of our result as well as to look at the effects of varying system parameters and conclude by noting some of the work that is being taken up.

II. FORMALISM

A. System definition

We consider a system consisting of one receiver and one transmitter with a wireless channel for communication. Moreover fading channel has been assumed. For a fading wireless channel, the maximum rate of information transfer i.e. capacity of the channel (due to Shannon [9]) is

$$C = \log(1 + SNR) \qquad SNR = \frac{Ph}{N_0 W}$$

here P is the transmitted power, h is the channel-fade coefficient and $N_0 W$ is the noise spectral density (SNR thus is the signal-to-noise ratio). The channel-fade coefficient, $h \in \mathcal{H} = \{e_1, e_2, \ldots, e_N\}$ according to the known probability distribution $P_H(\cdot)$. We assume a memoryless channel and $\mathcal{H}$ represents the set of possible channel states, where $e_i < e_j$ for $i < j$. On the transmitter side, power is provided by a rechargeable battery which has finite capacity to store energy (this could be the model for remote sensors placed in obscure areas which can be recharged periodically using only renewable sources

like wind and solar energy and which will have a limited capacity to store energy). Our main aim is to find the *optimal power allocation policy* for this system, which will tell us the rule by which power is to be used for data transmission in terms of the other parameters of the system so as to get maximum rate of information transfer. Time is considered to be slotted and we also assume full channel-side information (CSI). So we have perfect channel state information before transmission in every slot.

Let the energy in the battery at the beginning of the nth time slot be ξ_n and power allocated in the slot be P_n (energy per slot). We will use the random variable X_n to model the amount of recharging energy added to the battery at the end of nth slot by the renewable source. Note that the process $\{X_n\}_{n \geq 1}$ is assumed to be *i.i.d.* and random variable has a finite support in the set $\{0, 1, \ldots, a\}$. All our variables are over non-negative integers. (For example in Solar energy refer to [10] for the model relating to the exact distribution on X). Using these we can write our system equation

$$\xi_{n+1} = \min\left((\xi_n - P_n)^+ + X_n, \quad \xi_m\right) \qquad (1)$$

$(x)^+ = \max\{x, 0\}$ and here ξ_m is maximum energy that can be stored in the battery.

B. Markov Decision Process formulation

To solve this problem we are going to formulate it as an infinite horizon *Markov Decision Process* (MDP). The state space, S, will be two-dimensional, a typical state would be (ξ, h), which represents the current energy in the battery and the current channel-fade coefficient. From this the size of the state space will be $|S| = (\xi_m + 1) \times N$ (note that energy in the battery can be 0). Valid action space (power allocation) for the state (ξ, h) will be $P \in \{0, 1, \ldots, \xi\}$, this is because at any time we can at most allocate all the power available in the battery and also that we can also choose to allocate zero power (using this the $(\cdot)^+$ sign in the system equation becomes redundant). Union of all action spaces will be $A = \{0, 1, \ldots, \xi_m\}$. We will consider discounted rewards with a constant discount factor $\lambda \in (0, 1)$.

Our reward function $r : S \times A \to \mathcal{R}_0^+$ is

$$r\left((\xi, h), P\right) = \log\left(1 + \frac{hP}{N_0 W}\right)$$

Now we define **optimal reward function** $J^\star : S \to \mathcal{R}_0^+$ as the optimal value for each state that we start with. Transition Probability Matrix (TPM), $\left[\mathbb{P}\{(\xi_0, h_0) \mid (\xi, h), P\}\right]$, represents the probability of getting to some state (ξ_0, h_0) starting from (ξ, h) and taking action P.

Using all of the above we can write the Bellman's equation of dynamic programming as

$$J^\star(\xi, h) = \max_{P \leq \xi} \left\{ \log\left(1 + \frac{hP}{N_0 W}\right) + \right.$$

$$\left. \lambda \sum_{\xi_0 = (\xi - P)}^{\xi_m} \sum_{h_0 = 1}^{N} \mathbb{P}\{(\xi_0, h_0) \mid (\xi, h), P\} \times J^\star(\xi_0, h_0) \right\}$$

we will write this succinctly as (using $\mathbf{s} \equiv (\xi, h)$ as state)

$$J^\star(\mathbf{s}) = \max_{P \leq \xi} \left\{ r(\mathbf{s}, P) + \lambda \mathbb{E}_{h_0}^X\big(J^\star(f(\mathbf{s}, P), h_0))\big) \right\} \quad (2)$$

here f represents the rhs in (1).

Policy for this system will be map from state space to action space for each epoch, but as this is an infinite horizon MDP we will only look at **Stationary Deterministic** Policies to get the maximum throughput. So the optimal policy for our problem will be of the form $\pi^\star = \{\mu^\star, \mu^\star, \ldots\}$ and for convenience lets call it policy $\mu^\star$. So we can write the equation for optimal decision rule $\mu^\star : S \to A$ succinctly as follows

$$\mu^\star(\mathbf{s}) = \arg\max_{P \leq \xi} \left\{ r(\mathbf{s}, P) + \lambda \mathbb{E}_{h_0}^X\big(J^\star(f(\mathbf{s}, P), h_0))\big) \right\}.$$

With this our formulation of this problem is done and now we can move towards some of the results.

III. RESULTS

Here we prove structural results about monotonicity of $J^\star$ and $\mu^\star$ for our optimal power allocation problem which we have formulated as an MDP.

In the previous section we wrote the Bellman's Equation for our MDP and one way to solve it is using *Value Iteration* procedure (refer to the book by Bertsekas [11]). For this we start with an initial value (estimate) for the optimal reward function, say $J_0(\mathbf{s}) = 0 \ \forall \ \mathbf{s} \in S$ and then write iteration equations as

$$J_{k+1}(\mathbf{s}) = \max_{P \leq \xi} \left\{ r(\mathbf{s}, P) + \lambda \mathbb{E}_{h_0}^X\big(J_k(f(\mathbf{s}, P), h_0))\big) \right\} \quad (3)$$

where $\mathbf{s} = (\xi, h)$. From the theory of infinite horizon discounted reward MDP problems we know that this will converge (to $J^\star$) under the condition of bounded reward per stage (which is satisfied by the reward function in our case, the reward function is bounded and the action space and state space are all finite due to discrete nature of our formulation).

A. Preliminary Results

Here we will state and prove lemmas which will be required later to prove the main theorem.

Lemma 1 (Monotone Optimal Reward Function). *The optimal Reward Function, $J^\star(\xi, h)$, is non-decreasing in both arguments. We have two parts in this,*
  1) *For any $\xi \in \{0, \ldots \xi_m\}$,*

$$J^\star(\xi, h^+) \geq J^\star(\xi, h^-) \quad where \quad h^+ > h^-,$$

  2) *For any $h \in \mathcal{H}$,*

$$J^\star(\xi^+, h) \geq J^\star(\xi^-, h) \quad where \quad \xi^+ > \xi^-.$$

Proof: (*Part 1*) Take any ξ and consider channel states h^- and h^+ where $h^+ > h^-$. Notice that as the channel process is *i.i.d.*, the channel transitions are independent of each other. Specifically, we can say that the future channels are independent of current channel state, so the second term in (2) for $J^\star(\xi, h^+)$ and $J^\star(\xi, h^-)$ will be identical (as a function of P). Take $P^- = \mu^\star(\xi, h^-)$, by using (2) at this power we have

$$J^\star(\xi, h^+) - J^\star(\xi, h^-) \geq$$

$$\log\left(1 + \frac{h^+ P^-}{N_0 W}\right) - \log\left(1 + \frac{h^- P^-}{N_0 W}\right) \geq 0 \quad (4)$$

Proof: (*Part 2*) Take any h and consider ξ^+ and ξ^- where $\xi^+ > \xi^-$. Starting the value iteration with $J_0(\mathbf{s}) = 0 \ \forall \ \mathbf{s} \in S$ we will use induction to prove our result (for every step of value iteration). The base case is vacuously true. Now we assume that $J_k(\xi, h)$ is non-decreasing in ξ. Let P_k^- maximize the r.h.s of (3) for the state (ξ^-, h). From our iteration equations we have at power $P = P_k^-$ and for $D = J_{k+1}(\xi^+, h) - J_{k+1}(\xi^-, h)$

$$D \geq \lambda \mathbb{E}_{h_0}^X\big[J_k(f(\xi^+, P), h_0) - J_k(f(\xi^-, P), h_0)\big] \quad (5)$$

Since $\xi^+ > \xi^-$, then for the same power P_k^-, we'll have $f(\xi^+) > f(\xi^-)$ (for every instance of X). By induction hypothesis $J_k(\xi, h)$ is non-decreasing in ξ, hence the term inside the expectation in (5) is non-negative (for every instance of X and h). Hence after taking the expectation we will have

$$J_{k+1}(\xi^+, h) \geq J_{k+1}(\xi^-, h)$$

using induction now we can claim the above $\forall \ k \in \mathbb{Z}^+$ and hence the result follows by taking $\lim_{k \to \infty}$. ∎

The above lemma can be effectively written as

$$J^\star(\xi^+, h^+) \geq J^\star(\xi^-, h^-) \quad \forall \ \xi^+ \geq \xi^-, \ h^+ \geq h^-$$

Now that we have shown monotonically increasing nature of optimal reward function, another property that will go a long way in proving our final result is that

of concavity of $J^\star$. Typically concavity (convexity) and equivalently sub-modularity (super-modularity) has been the most used method to prove monotonicity of policy. So here with the help of a little extra set up we prove the important property of concavity of $J^\star$ in energy only.

Lemma 2 (Concave Optimal Reward Function). *The optimal reward function $J^\star(\xi, h)$ is **concave** in ξ for a fixed h.*

Proof: Here we will use induction on Value iteration steps, just like before. We will first show that concavity in J_k implies concavity in J_{k+1}. Assuming J_k is concave we take states as

$$s_1 = (\xi_1, h) \quad s_2 = (\xi_2, h) \quad \bar{s} = (\xi, h)$$

where $\xi = \alpha\xi_1 + (1-\alpha)\xi_2 \quad (0 < \alpha < 1)$. Now taking the optimal powers for this step of the iteration as P_1 and P_2 we can write the equations

$$J_{k+1}(s_1) = r(s_1, P_1) + \lambda \mathbb{E}_{h_0}^X \big[J_k(f(s_1, P_1), h_0) \big]$$
$$J_{k+1}(s_2) = r(s_2, P_2) + \lambda \mathbb{E}_{h_0}^X \big[J_k(f(s_2, P_2), h_0) \big]$$

We know that $\log(\cdot)$ reward here is a concave function in P and is constant w.r.t variation in ξ, hence we have

$$\alpha r(s_1, P_1) + (1-\alpha)r(s_2, P_2) \leq r\left(\bar{s}, \bar{P}\right) \qquad (6)$$

where $\bar{P} = \alpha P_1 + (1-\alpha)P_2$ and $\bar{s}$ can be used because it has the same channel coefficient, h, as s_1 and s_2 . By induction hypothesis J_k is concave as well, so

$$\alpha J_k(f(s_1, P_1), h_0) + (1-\alpha)J_k(f(s_2, P_2), h_0)$$
$$\leq J_k\big(\alpha f(s_1, P_1) + (1-\alpha)f(s_2, P_2), h_0\big) \qquad (7)$$

Beyond this point we divide the problem into cases, depending on the values of X.

Case 1: All X, such that $f(s_1, P_1), f(s_2, P_2) < \xi_m$.

$$\Rightarrow \alpha f(s_1, P_1) + (1-\alpha)f(s_2, P_2)$$
$$= \alpha\xi_1 + (1-\alpha)\xi_2 - (\alpha P_1 + (1-\alpha)P_2) + X$$
$$= \xi - \bar{P} + X = f\left(\bar{s}, \bar{P}\right)$$

The last equality follows since the argument in this case is clearly $< \xi_m$. Hence continuing from (7) we can write

$$\alpha J_k(f(s_1, P_1), h_0) + (1-\alpha)J_k(f(s_2, P_2), h_0)$$
$$\leq J_k\left(f\left(\bar{s}, \bar{P}\right), h_0\right) \qquad (8)$$

Using (6) and (8) we can thus write

$$\alpha J_{k+1}(s_1) + (1-\alpha)J_{k+1}(s_2)$$
$$\leq r\left(\bar{s}, \bar{P}\right) + \lambda J_k\left(f\left(\bar{s}, \bar{P}\right), h_0\right) \qquad (9)$$

Case 2: All X, such that $f(s_1, P_1) = \xi_m = f(s_2, P_2)$.

$$\Rightarrow \alpha(\xi_1 - P_1 + X) + (1-\alpha)(\xi_2 - P_2 + X) \geq \xi_m$$

so $f(\bar{s}, \bar{P}) = \xi_m$ and hence we can write

$$J_k\big(\alpha f(s_1, P_1) + (1-\alpha)f(s_2, P_2), h_0\big)$$
$$= J_k\left(\xi_m, h_0\right) = J_k\left(f(\bar{s}, \bar{P}), h_0\right)$$

from this the same result as in (9) follows.

Case 3: All X, such that $f(s_2, P_2) < f(s_1, P_1) = \xi_m$.

$$\Rightarrow \xi_2 - P_2 + X < \xi_m = \xi_1 - P_1 + X - \beta \ (\beta \geq 0),$$
$$\alpha f(s_1, P_1) + (1-\alpha)f(s_2, P_2) = \xi - \bar{P} + X - \alpha\beta \quad (10)$$

Clearly the term in the r.h.s in (10) is less than ξ_m and it also is $\leq \left(\xi - \bar{P} + X\right)$ so we can conclude

$$\xi - \bar{P} + X - \alpha\beta \leq \min\{\xi - \bar{P} + X, \ \xi_m\}$$

Since J_k is non-decreasing in energy (shown in the proof of Lemma 1) we can conclude the same as in (8) and from there (9) as well. Cases finished.

From these three cases what we have seen that (9) is satisfied for all h_0 and all possible values of X and hence we can introduce the $\mathbb{E}(\cdot)$ operator and conclude

$$\alpha J_{k+1}(s_1) + (1-\alpha)J_{k+1}(s_2) \leq r\left(\bar{s}, \bar{P}\right) +$$
$$\lambda \mathbb{E}_{h_0}^X \left[J_k\left(f\left(\bar{s}, \bar{P}\right), h_0\right) \right] \leq J_{k+1}(\bar{s})$$

where the last inequality holds because $\bar{P}$ can generate a value only less that or equal to the optimal value for state $\bar{s}$ (at the $(k+1)^{th}$ iteration).

Now from all this we have shown that concavity in J_k implies concavity in J_{k+1} and starting with a concave initial value of the iteration like $J_0(s) = 0 \ \forall \ s \in S$, we can conclude by induction that J_k is concave in ξ $\forall \ k \in \mathbb{Z}^+$. Hence as Value iteration converges we can conclude that $J^\star$ is concave in ξ. ∎

Corollary 1. *If we have energy levels $x \leq w \leq z \leq y$ such that*

$$x + y = w + z \qquad then \qquad (11)$$
$$J^\star(x, h) + J^\star(y, h) \leq J^\star(w, h) + J^\star(z, h)$$

Proof: For a fixed h define $J^\star(\xi, h) \equiv g(\xi)$. Also let $\Delta g(i) = g(i+1) - g(i)$, then we can write

$$g(x, h) + g(y, h) = 2g(x) + \sum_{i=x}^{y-1} \Delta g(i)$$

$$g(w, h) + g(z, h) = 2g(x) + \sum_{i=x}^{w-1} \Delta g(i) + \sum_{i=x}^{z-1} \Delta g(i)$$

As $J^\star$ is concave in energy, we know that $\Delta g(i)$ is non-increasing with i (following the "Law of diminishing returns" for concave functions). Summations in both equations above have the same number of terms (due to (11)) and clearly the first equation sums $\Delta g(i)$ over higher values of i and therefore is smaller. ∎

This property is called *sub-modularity*.

B. Main Structural Result

Now we prove the main structural result with the aid of the lemmas of previous subsection.

Theorem 1 (Monotonic Optimal Policy). *The optimal policy of power allocation, $\mu^\star(\xi, h)$, is non-decreasing in both arguments. We have two parts in this,*

1) *For any $\xi \in \{0, \ldots \xi_m\}$,*

$$\mu^\star(\xi, h^+) \geq \mu^\star(\xi, h^-) \quad \text{where} \quad h^+ > h^-$$

2) *For any $h \in \mathcal{H}$,*

$$\mu^\star(\xi^+, h) \geq \mu^\star(\xi^-, h) \quad \text{where} \quad \xi^+ > \xi^-$$

Proof: (*Part 1*) Consider two channel states h^- and h^+ where $h^+ > h^-$. We can write

$$\mu^\star(\xi, h^+) = arg \max_{P \leq \xi} \left\{ \log\left(1 + \frac{h^+ P}{N_0 W}\right) - \log\left(1 + \frac{h^- P}{N_0 W}\right) \right.$$
$$\left. + \log\left(1 + \frac{h^- P}{N_0 W}\right) + \lambda \mathbb{E}_{h_0}^B \left[J^\star(f(\xi, P), h_0) \right] \right\}$$

Since the last term is independent of h^+ we have $\mu^\star(\xi, h^+) = \max_{P \leq \xi} \{T_1 + T_2\}$ where

$$T_1 = \log\left(1 + \frac{h^+ P}{N_0 W}\right) - \log\left(1 + \frac{h^- P}{N_0 W}\right)$$

and T_2 is the full term that will appear inside the $\max$ operator in the expression for $\mu^\star(\xi, h^-)$, which means that T_2 achieves its maximum at $P_{h^-} = \mu^\star(\xi, h^-)$. Notice that T_1 is monotonically increasing in P, since

$$\frac{dT_1}{dP} = \frac{N_0 W(h^+ - h^-)}{(N_0 W + h^+ P)(N_0 W + h^- P)} > 0 \quad (12)$$

Considered at any $P < P_{h^-}$, the term T_1 will have a value lesser than at P_{h^-} (because its monotonically increasing) and same for T_2 (because maxima is at P_{h^-}). Hence $\{T1 + T2\}$ cannot achieve its maxima for any $P < P_{h^-}$ and we conclude

$$\mu^\star\left(\xi, h^+\right) \geq \mu^\star(\xi, h^-)$$

Proof: (*Part 2*) Firstly note that

$$\xi_2 < \xi_m \Rightarrow \mathbb{P}\{\xi_2 \mid \xi, P\} = \mathbb{P}\{X = \xi_2 - \xi + P\}$$
$$\xi_2 = \xi_m \Rightarrow \mathbb{P}\{\xi_m \mid \xi, P\} = \mathbb{P}\{X \geq \xi_m - \xi + P\}$$

From the above now we can write the second term in $J^\star$ as

$$\sum_{\xi_0 = \xi - P}^{\xi_m} \sum_{h_0 = 1}^{N} \mathbb{P}\{h_0\} \times \mathbb{P}\{\xi_0 \mid \xi, P\} \times J^\star(\xi_0, h_0) \equiv$$

$$\mathbb{E}_{h_0} \left[\sum_{i=0}^{\xi_m - \xi + P - 1} \mathbb{P}\{X = i\} \times J^\star(\xi - P + i, h_0) + \right.$$

$$\left. \mathbb{P}\{X \geq \xi_m - \xi + P\} \times J^\star(\xi_m, h_0) \right]$$

but we can write $P(X \geq \xi_m - \xi + P)$ in terms of the summation preceding it, hence we will have

$$J^\star(\xi, h) = \lambda \mathbb{E}_{h_0} \left[J^\star(\xi_m, h_0) \right] +$$
$$\max_{P \leq \xi} \left\{ \log\left(1 + \frac{Ph}{N_0 W}\right) - \lambda \mathbb{E}_{h_0} \sum_{i=0}^{\xi_m - \xi + P - 1} \mathbb{P}\{X = i\} \right.$$
$$\left. \times \left[J^\star(\xi_m, h_0) - J^\star(\xi - P + i, h_0) \right] \right\} \quad (13)$$

Now we will use contradiction to prove our result i.e. assume that there exists states $\xi_1 > \xi_2$ with optimal powers $P_1 < P_2$.

Let $J_P(\xi, h)$ represents the rhs term in (2), evaluated at power P. Then due to optimality of P_2 with ξ_2 and P_1 with ξ_1 we will have the equations

$$J_{P_2}(\xi_2, h) - J_{P_1}(\xi_2, h) \geq 0,$$
$$J_{P_1}(\xi_1, h) - J_{P_2}(\xi_1, h) \geq 0$$

Adding the two equations with the help of (13) and using $g(\xi) \equiv J^\star(\xi, h)$ as well as $p_i \equiv \mathbb{P}\{X = i\}$ will give us

$$\mathbb{E}_h \left[\sum_{i=0}^{\kappa_{11}} p_i A(i) + \sum_{i=\kappa_{11}+1}^{\kappa_{12}} p_i B(i) + \right.$$
$$\left. \sum_{i=\kappa_{12}+1}^{\kappa_{21}} p_i C(i) + \sum_{i=\kappa_{21}+1}^{\kappa_{22}} p_i D(i) \right] \geq 0 \quad (14)$$

for $\kappa_{ij} = \xi_m - y_{ij} - 1$, $y_{ij} = (\xi_i - P_j)$ $i, j \in \{1, 2\}$
$A(i) = g(y_{11} + i) + g(y_{22} + i) - g(y_{12} + i) - g(y_{21} + i)$,
$B(i) = g(\xi_m) + g(y_{22} + i) - g(y_{12} + i) - g(y_{21} + i)$,
$C(i) = g(y_{22} + i) - g(y_{21} + i)$,
$D(i) = -g(\xi_m) + g(y_{22} + i)$.

In breaking the above summations appropriately we have assumed w.l.o.g. $\kappa_{12} \leq \kappa_{21}$, which means $\kappa_{11} \leq \kappa_{12} \leq \kappa_{21} \leq \kappa_{22}$ & $y_{11} \geq y_{12} \geq y_{21} \geq y_{22}$.

We will argue that (14) is a contradiction. Our following calculations hold for every h.

Simply by our construction $y_{22} \leq y_{12}, y_{21} \leq y_{11}$ and

$$y_{11} + y_{22} = (\xi_1 + \xi_2) - (P_1 + P_2) = y_{21} + y_{12}$$

so by Corollary 1, $A(i) \leq 0 \ \forall \ i$. We know that g is non-decreasing (Lemma 1). As $y_{22} \geq y_{21}$ we'll have $C(i) \leq 0 \ \forall \ i$. Since the range of summation for $D(i)$ is such that $y_{22} + i \leq \xi_m$ we also have $D(i) \leq 0 \ \forall \ i$.

Now looking at $B(i)$, define successive differences $\Delta g(l) = g(l + 1) - g(l)$ (using the same method as in Corollary 1). Due to concavity of $J^\star$ (Lemma 2) this is non-increasing. We can express $g(\xi_m), g(y_{12} + i)$ and $g(y_{21} + i)$ as a summation of Δg starting from $g(y_{22} + i)$. We will then see here that $g(\xi_m) + g(y_{22} + i)$ has fewer

Δg terms in summation compared to $g(y_{12}+i)+g(y_{21}+i)$ and those $\Delta g(l)$ terms are also smaller since they are being summed over higher l. Since Δg is positive we can conclude that $B(i) \leq 0 \; \forall \; i$.

So from all this we have shown that all terms in (14) are negative $\forall \; h$ and thus when their expectation is taken, it will be negative too. Thus we have shown a contradiction. Hence proved. ∎

The above result can be concisely written as

$$\mu^\star(\xi^+, h^+) \geq \mu^\star(\xi^-, h^-) \quad \forall \; \xi^+ \geq \xi^-, \;\; h^+ \geq h^-$$

C. Possible Generalizations

In this problem we had compact support on X and ξ. Note that as long as we have compact support for these two, the results will carry through to uncountable state/action space as well. Meaning, instead of having discrete values of ξ and X, we can make it continuous (over real numbers) and end up with the same results.

The reward function used here was $\log$, we can enlist the following properties that were used explicitly in proving our results

1) reward (r) depends only on h and P, its independent of ξ (used in Lemma 1 part 1),
2) $r((\xi,h),P)$ is **concave** in P (used in Lemma 2),
3)
$$\frac{\partial r((\xi,h),P)}{\partial h} \geq 0 \quad \text{(used in (4))}.$$
4)
$$\frac{\partial^2 r((\xi,h),P)}{\partial P \partial h} \geq 0 \quad \text{(used in (12))}.$$

No other property of $\log$ function was used. This means that any reward function satisfying these three properties will give us the same results. (Reward function is assumed to be positive for all state/action pairs)

IV. SIMULATION RESULTS

We present here simulation results which essentially verify our results (the properties proved here were verified for a large number of parameters before being proved).

We take the parameters in the problem as

$$\xi_m = 50 \qquad a = 56 \qquad \lambda = 0.85 \qquad N = 17$$

and $N_0 W = 10$. This means that the channel states are in $\mathcal{H} = \{1, \ldots, 17\}$. The distribution h is taken to be *bell-shaped* and distribution on X was taken to be a strictly decreasing one. For this system we first plot the optimal policy $\mu^\star(\xi,h)$, (which we have proved to be non-decreasing in both ξ and h),
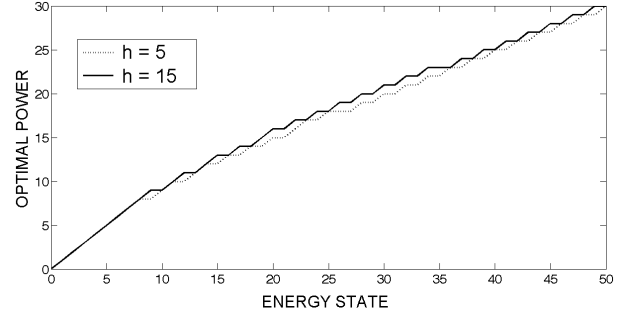

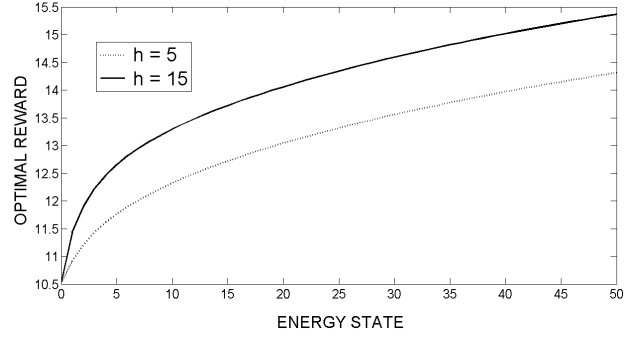Fig. 1. $\mu^\star(\xi,h)$ vs. ξ for $h = 5, 15$

Fig. 2. $J^\star(\xi,h)$ vs. ξ for $h = 5, 10$

and then the optimal reward function $J^\star(\xi,h)$, which should not only be non-decreasing in both arguments but also concave in ξ.

Apart from verifying our proven results another important feature to discuss is the structure of the random power being added in every slot i.e. distribution of X. Higher power added in every slot should give us higher optimal powers to work with, since even if we spend power on a bad channel once, we wouldn't have to wait long before the battery gets recharged (since higher values of X are more likely). In this regard we also present here the graph of $\mu^\star$ for 2 different distributions on X. $\mathbb{P}_{X_1}$ represents a distribution which decreases with x (this is also the distribution we have been using till now) and $\mathbb{P}_{X_2}$ represents a distribution which is exactly inverted i.e. it increases with x. Clearly $\mathbb{P}_{X_2}$ has higher mean that $\mathbb{P}_{X_1}$.

As an instructive example we can also look at the solution after varying λ, variation in λ is of central importance because it essentially tells us how much importance is being given to future rewards as opposed to the current reward, which basically dictates the average number of recharge cycles that the battery may have to go through (and consequently its effective life-time).

We notice in our case that as λ increases more importance is given to future rewards and consequently optimal powers become lower i.e. power is being saved for future
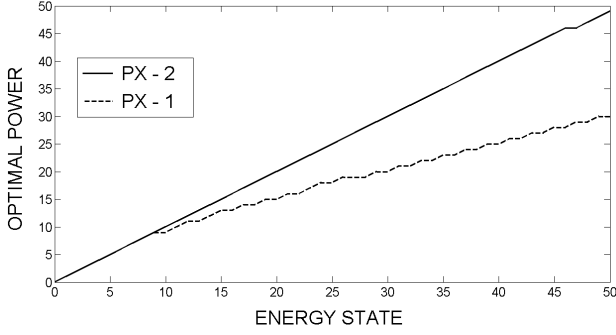
Fig. 3. $\mu^\star(\xi, h)$ vs. ξ for $\mathbb{P}_{X_1}$, $\mathbb{P}_{X_2}$ and $h = 10$

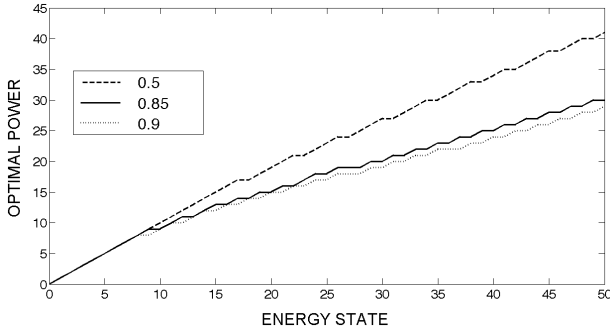where probably better channels may be available.

Fig. 4. $\mu^\star(\xi, h)$ vs. ξ for $\lambda = 0.5, 0.85, 0.9$ and $h = 15$

V. CONCLUSION

In this paper we have proved one of the most important features of the power allocation problem constrained under limited capacity of the battery. The results have been proved from scratch without the use of any known results except the standard ones for a general MDP setting. The most pleasing aspect of this result is that there were no assumptions required on the distribution of X and h, just that their respective processes are $i.i.d.$. Along with the main result, the side results like the monotone and concave nature of $J^\star$ are also important tools in deciding a minimum complexity algorithm.

Once we have a monotonically increasing optimal policy then not only does the search space for any algorithm gets reduced but also the memory required to store the related tables gets reduced, which is very much desirable as the sensors are quite small in size. The policy here is an off-line policy.

The other results being looked into are that of finding an actual algorithm that will take full advantage of the results proved here. Further work that is going on is for the case of unknown channel process, in which case Q-learning methods need to be looked into and

possibly an on-line policy can be determined. Another possibility is that of $\{X_n\}_{n \geq 1}$ process being dependent on state, which actually is a realistic scenario in capacitor charging models given for solar cells.

REFERENCES

[1] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *Networking, IEEE/ACM Transactions on*, vol. 12, no. 4, pp. 609 – 619, aug. 2004.

[2] Y. T. Hou, Y. Shi, and H. D. Sherali, "Rate allocation in wireless sensor networks with network lifetime requirement," in *Proceedings*, ser. MobiHoc '04, 2004, pp. 67–77.

[3] Y. Ammar, A. Buhrig, M. Marzencki, B. Charlot, S. Basrour, K. Matou, and M. Renaudin, "Wireless sensor network node with asynchronous architecture and vibration harvesting micro power generator," in *Proceedings*, ser. sOc-EUSAI '05, 2005, pp. 287–292.

[4] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljai, "Wireless power transfer via strongly coupled magnetic resonances," vol. 317, no. 5834, pp. 83–86, 2007.

[5] R. F. Serfozo, "Monotone optimal policies for markov decision processes," in *Stochastic Systems: Modeling, Identification and Optimization, II*, ser. Mathematical Programming Studies.

[6] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1994.

[7] Z. Mao, C. Koksal, and N. Shroff, "Resource allocation in sensor networks with renewable energy," in *Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on*, aug. 2010, pp. 1 –6.

[8] S. Chen, P. Sinha, N. Shroff, and C. Joo, "Finite-horizon energy allocation and routing scheme in rechargeable sensor networks," in *Proceedings of IEEE INFOCOM, 2011*, april 2011, pp. 2273 –2281.

[9] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, pp. 3–55, January 2001.

[10] C. Renner, J. Jessen, and V. Turau, "Lifetime prediction for supercapacitor-powered wireless sensor nodes."

[11] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Two Volume Set*, 2nd ed. Athena Scientific, 2001.